

## ソーシャルブックマークサービスを利用した 情報レコメンデーション

○白土 慧 (北大) 吉井 伸一郎 (北大) 古川 正志 (北大)

本論文では、ユーザが編纂した情報に基づく情報推薦システムの開発を目的とする。個人によって編纂された幅広い情報を用いた推薦を行うために、ソーシャルブックマークサービスを対象とする。適切な情報を抽出するため、Web 上から各ユーザのデータを収集、ユーザ関連度を解析するシステムを構築した。興味の類似した複数のユーザが相互に関連する情報を持つと仮定し、ユーザ間の情報共参照性をネットワークとして表現する。ここから各ユーザを中心とする部分グラフを抽出することで情報を推奨する。さらに本論文では、指定ユーザと類似した嗜好を持つユーザの関連性をグラフとして表示し推薦情報を提示する Web システムを構築し、提案手法の有効性について検討する。

### Information recommendation using social bookmark service

Kei SHIRATSUCHI, Hokkaido University

Shinichiro YOSHII, Hokkaido University

Masashi FURUKAWA, Hokkaido University

In this paper, we develop the interested information recommendation system based on the user's compiled information. This system targets Social Bookmark Service (SBM). For finding appropriate interested information, we collected each user's bookmark entry on Web, and developed system which analyze user's association. We assume that users to whom interest is similar have associated information mutually, we describe user's co-citation information as network. We find interested information and subgraph that centers on each user from this network. And we develop web service to recommend new interested information. That service has list of the user who has similar preference to specified user, subgraph to the user, System to present recommended information to user.

#### 1. はじめに

近年 Web 上には、カスタマイズ性の高いツールを用いて自分に有益な情報を編纂するサービスが増加しており、個人による情報の編纂とそれに伴うコミュニケーションが活発に行われて

いる。

現在 Web 上では、Google に代表されるキーワードの組み合わせによる情報検索と、ニュースやブログなど特定 Web ページの RSS 等を利用した定期的な閲覧が主な情報収集の手段となっている。しかし、検索に用いるキーワード

はユーザの知識によって限定され、Web ページの定期的な閲覧で得られる情報はコンテンツ製作者の知識に限定される。そのため新たな情報の発見は難しく、ユーザの持つ知識を超えるような幅広い情報の発見は難しい。

本論文では、ユーザが編集した情報に基づいた情報推薦システムの開発を目的とする。対象はソーシャルブックマークサービスとする。このサービス上では個人によって編集された情報が集積されることで、幅広い情報を得ることができる。このサービスを利用したレコメンデーションによって、ユーザが未知の知識を発見できると考える。各ユーザに対して適切な情報を抽出するため、はてなブックマーク<sup>\*1</sup>を対象に Web 上から各ユーザのブックマークを収集、ユーザ関連度を解析するシステムを構築する。このシステムは、得られたデータからネットワークを構成し、リンクが密に存在する部分をコミュニティとして抽出する。さらに、解析したデータから情報を推薦を行うための Web サービスを開発する。

## 2. ソーシャルブックマークサービス

ソーシャルブックマークサービスは、これまで個人の PC 内に保存していたブックマークを Web 上に公開するサービスである。ユーザは、任意の URL をブックマークとしてサービス上に登録でき、これらにタグと呼ばれるキーワードを関連付けることができる。サービスに登録された URL はエントリと呼ばれる。他人のブックマークを閲覧することでこれを有益な情報源としたり、自分の趣味や興味のある事柄についての URL を自ら登録することで、自身の関心を容易に公開できる。ソーシャルブックマークについての研究としては、

表 1 ブックマークデータの例

ユーザ	登録したエントリ
ユーザ A	エントリ 1 エントリ 2 エントリ 3 エントリ 4
ユーザ B	エントリ 2
ユーザ C	エントリ 1 エントリ 2
ユーザ D	エントリ 1 エントリ 3
ユーザ E	エントリ 3 エントリ 4

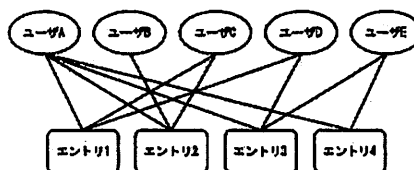


図 1 ブックマークデータの 2 部グラフでの構成

del.icio.us<sup>\*2</sup>の解析を行ったものがある [1].

## 3. 二部グラフによるモデル化

得られたデータに対して、ユーザ、エントリをそれぞれノードとし、あるユーザとそのユーザが登録したエントリ間にリンクを張ることでネットワークを構成する。このネットワークを  $G$  とおくと、 $G$  におけるノード集合はユーザとエントリによる 2 つの素な集合に分割され、この集合をそれぞれ  $U, E$  とおくと、 $G$  のすべてのリンクは  $U$  のノードと  $E$  のノードを結ぶ。つまり  $G$  は二部グラフとなる。

例として、表 1 のようなデータの場合、図 1 のように表すことができる。

<sup>\*1</sup> <http://b.hatena.ne.jp/>

<sup>\*2</sup> <http://del.icio.us/>

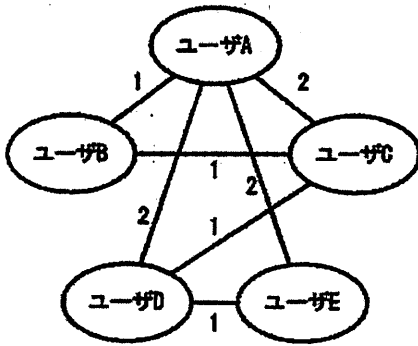


図2 ブックマークデータの1部グラフでの構成

#### 4. コミュニティ抽出

本研究ではエントリの共参照関係からネットワークを構成し、ユーザのみをノードとする1部グラフとして扱う。これは、ユーザ数に対してエントリ数が膨大になることが予想されるため、解析における計算量を減少させることを目的とした。共参照関係によるネットワークとは、ある2人のユーザが同一のエントリをブックマークしているとき、そのユーザ間にリンクを張ることで構成するものである。

表1のデータにおける共参照関係ネットワークを図2に示す。リンクの数字は、そのユーザ間の共参照したエントリの数である。

本研究では、Clauset[2]による *Local Modularity* によるコミュニティ抽出方法を採用した。さらに、リンクの重みを考慮して、拡張となる手法を提案しこれを実装した。

まず、共参照関係ネットワークでは、ある2ユーザが同一のエントリをブックマークしているとそのユーザ間にリンクが張られるが、このとき複数のエントリを同様にブックマークしていることが考えられる。このときの共通エ

ントリ数をリンクの重みとして表現することができる(図2)。重みつきネットワークに関しては、Newmanによる研究がある[3]。この中でNewmanは、 $n$ ノードの重みなしネットワークの $n \times n$ 隣接行列を

$$A_{ij} = \begin{cases} 1 & \text{ノード } i, j \text{ が接続されている} \\ 0 & \text{それ以外} \end{cases}$$

としたとき、重みつきネットワークにおける隣接行列を以下のように定義した。

$$A_{ij} = \text{ノード } i \text{ から } j \text{ へのリンクの重み}$$

本研究において提案するアルゴリズムを紹介する。グラフ全体を $G$ とし既にコミュニティとした部分集合を $C$ 、 $C$ に含まれず $C$ とリンクがある、つまり隣接している部分集合を $U$ とする。さらに、 $C$ 中で $U$ と隣接している部分集合を $B$ とおく。

$B$ の隣接行列を以下のようにおく。

$$B_{ij} = \text{ノード } i, j \text{ が接続されており、} \\ \text{どちらかが } B \text{ に属するときの} \\ \text{リンクの重み}$$

*Local Modularity* を $R$ とおくと、以下のよう

$$R = \frac{\sum_{ij} B_{ij} \delta(i, j)}{\sum_{ij} B_{ij}} = \frac{I}{T} \quad (1)$$

ここで $\delta(i, j)$ は、 $v_i \in B$ かつ $v_j \in C$ 、もしくは $v_i \in C$ かつ $v_j \in B$ のとき1、それ以外の場合0であるとする。つまり、 $T$ は $B$ の総リンク数、 $I$ はそのうち $U$ でないノードとつながっているリンクの和となる。

本手法におけるコミュニティは、1つのシードノードから始め、*Local Modularity* を最大にするように $C$ を構成することで求められる。このアルゴリズムは以下ようになる。

1. シードノード  $v_0$  を  $C$  に追加し  $B$  を  $v_0$  とする。さらに  $v_0$  の隣接ノード集合を  $U$  に追加。
2. 以下の手順を  $C$  のノード数が決められた大きさになるまで繰り返す。
3.  $v_j \in U$  に対して  $\Delta R_j$  を計算する
4.  $\Delta R_j$  が最大となる  $v_j$  を  $C$  に追加し、 $v_j$  の隣接ノードのうち新規に現れたものを  $U$  に追加する。
5.  $R$  と  $B$  を更新する。

ここで、 $v_j$  に対する  $\Delta R_j$  は式 (1) から導かれる以下の式を利用して高速に計算できる。

$$\Delta R_j = \frac{x - Ry - z(1 - R)}{T - z + y} \quad (2)$$

式 (2) における  $x, y, z$  は以下ようになる。

- $x$  は  $v_j$  の  $B$  へのリンクの重みの和
- $y$  は  $v_j$  が  $C$  へ選ばれたとき  $T$  に加わるリンクの重みの和
- $z$  は  $v_j$  が  $C$  へ選ばれたとき  $T$  から減るリンクの重みの和

## 5. 提案システム

ソーシャルブックマークサービスを利用して、ユーザの興味に基づいた情報レコメンデーションを行う。これは、興味の近いユーザを発見しユーザ同士で未知の情報を補い合うことで行うとし、実際には以下の手順をとる。

- データを収集し、ユーザをノードとした共参照関係ネットワークを構成する。
- このネットワークから、推薦対象となるユーザを中心としてリンクが密につながった部分をコミュニティとして抽出する。
- コミュニティとして抽出されたユーザは関心が近いとして、Java アプレットを利用してこれらのユーザ間の部分グラフ、ユーザがブックマークに登録していないエントリを推薦する情報として提示する。

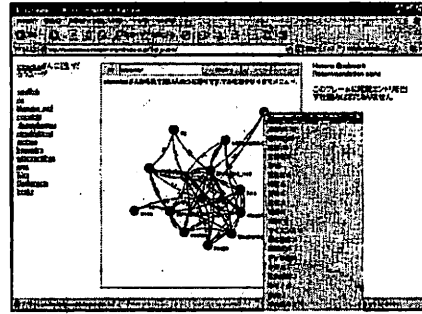


図3 結果表示システムを用いて表示したコミュニティ構造の例

## 6. 解析結果

はてなブックマークより 2005 年 11 月 22 日から 2006 年 1 月 1 日の期間に幅優先探索を用いたクローリングによってデータを取得した。このデータは、ユーザ数 7,302, エントリ数 792,993, ブックマーク数 1,745,986 であった。

全ユーザ 7,302 人に対し、それぞれ提案手法を適用してコミュニティ抽出を行う。このときコミュニティの最大ノード数を 21 と設定した。つまり、関連のあるユーザとして最大で 20 人を選び出すということである。

コミュニティの解析結果として、今、あるユーザ（ここでは TARGET\_USER とする）に対する関連推奨情報を有するユーザからなるコミュニティ構造を図 3 に示す。中心の Java アプレットに表示されたグラフが推薦対象ユーザを中心としたコミュニティ構造であり、緑が推薦対象ユーザ、赤が関連度が高いユーザとなり、その間のリンクとその重みを表示している。左に関連度の高いユーザの一覧を表示している。推薦対象ユーザ TARGET\_USER と関連があるとして抽出されたコミュニティに含ま

表2 推薦対象ユーザ TARGET\_USER と本研究により得たコミュニティ内のユーザのタグの比較

ユーザ名	[タグ] 出現数
TARGET_USER	[Music]279 [Web]250 [iPod]171 [Blog]142 [Funny]138 [Business]120 [2ch]115 [Google]107 [Mac]103 [Media]72 [Law]71 [Game]71 [Korea]66 [Misc]62
USER_A	[massmedia]62 [avex]46 [life]42 [hatena]39 [nuts]37 [music]31 [korea]29 [tech]29 [misc]22 [2ch]21 [sony]20 [crime]18 [blog]18 [emc]14 [government]14
USER_B	[マスコミ]35 [偏向報道]14 [ソニー]13 [メディアリテラシー]11 [印象操作]11 [朝日]9 [靖国]8 [情報操作]7 [特定アジア]7 [情報]5 [avex]5 [のまねこ]5 [2ch]5
USER_C	[著作権]168 [CCCD]131 [It's a Sony]105 [私的録音録画補償金]101 [iTMS]99 [JASRAC]79 [音楽]69 [選挙]55 [音楽配信]44 [映画]42 [図書館]34 [法制小委]34
USER_D	[著作権]401 [図書館]208 [出版]167 [音楽]123 [文化審議会]121 [文字・活字文化振興法]90 [BLOG]82 [音楽配信]77 [電子書籍]64 [JASRAC]63 [web]56 [Google]56 [国会]54
USER_E	[業界]156 [テレビ]124 [ラテンアメリカ]107 [音楽配信]80 [コンテンツ]77 [ネット]69 [映画]67 [映像配信]64 [PVR]57 [音楽]51 [政策]43 [総選挙 2005]34 [著作権]31
USER_F	[music]295 [copyright]125 [book]72 [news]72 [column]55 [artist]47 [video]37 [音楽配信]37 [cd]37 [interview]32 [P2P]29 [site]29 [cccd]24 [audio]22 [批評]21
USER_G	[音楽配信]73 [論文]53 [著作権]44 [グルメ]21 [P2P]21 [鎌倉]10 [sns]9 [物欲]8 [音楽]7 [GTD]7 [Podcast]5 [fishmans]5 [Computer Music]5 [CCCD]4 [教育]4 [中古レコード]3
USER_H	[JASRAC]47 [著作権]15 [書籍]10 [公貸権]8 [SF]8 [図書館]6 [映画]5 [野球]5 [書店]5 [競馬]4 [SONY]4 [サイエンス]4 [新古書店]4 [補償金]4 [アメリカ]4 [ネタ]3 [私的複製]3
USER_I	タグの付加なし
USER_J	タグの付加なし
USER_K	[IPv6]47 [コンテンツ流通]34 [Skype]13 [RFID]7 [P2P]7 [VoIP]5 [WiMAX]4 [OSS]4 [ユニバーサルサービス]3 [wireless]2 [著作権]1 [住基ネット]1 [知財]1 [Archive]1 [電力線]1
USER_L	タグの付加なし
USER_M	タグの付加なし

れるユーザは 14 人であった。表 2 は推薦対象ユーザ TARGET\_USER と関連があるとして抽出されたコミュニティに含まれるユーザがエントリに対して付加したタグの一覧である。括弧内の単語がタグ、数字はそのユーザが付加した回数である。

## 7. 考察

ユーザがブックマークに付加するタグは、ユーザがエントリからどのような意味を抽出したかを表現する。よって、ユーザごとのタグの集合はそのユーザの興味を示すものと考えられ、推薦対象となるユーザと本研究で得たコミュニティでのユーザ集合では、意味の似たタグが見られると予想される。

表 2 を見ると、推薦対象であるユーザ TARGET\_USER は [Music] や [Web], [iPod] などというタグを頻繁につけており音楽やインターネットに興味があることがうかがえる。一方、得られたコミュニティに含まれるユーザのタグには [音楽] や [音楽配信], [music] などが見られた。そのほかにも推薦対象ユーザのタグ [Media] に対して、コミュニティ内のユーザから [業界] や [メディアリテラシー], [massmedia] など関連性の高いタグが見つけれられる。[著作権] や [copyright] などは推薦対象ユーザのタグには見られないが、実際に調べたところコミュニティ内のユーザが [著作権] などとタグをつけているエントリに対して [Music] と [Law] を付加していることがわかった。

この結果から、推薦対象となるユーザと本研究で得たユーザ集合との関心分野に一致が見られ、この集合による推薦対象ユーザへの未知情報の推薦が妥当であるといえる。

## 8. おわりに

本研究では、ユーザの編纂した情報に基づいて未知の関心情報を推薦するシステムの開発を行った。ユーザの編纂した情報としてソーシャルブックマークサービスを対象とし、Web 上から各ユーザのブックマークしたエントリを収集したのち、ユーザの関連度を解析するシステムを構築した。さらに、ユーザにとって未知の関心情報を推薦するための Web サービスを開発した。

今後の展望として、今回得たデータをさらに詳細に解析するとともに、継続的なデータ収集と可視化システムの改良を考えている。

## 文献

- [1] Shen Kaikai and Wu Lide. Folksonomy as a complex network, Sep 2005.
- [2] Aaron Clauset. Finding local community structure in networks. *Physical Review E*, Vol. 72, p. 026132, 2005.
- [3] M. E. J. Newman. Analysis of weighted networks, July 2004.